ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 4, Nomor 1, Tahun 2015, Halaman 11 - 20
Online di: http://ejournal-s1.undip.ac.id/index.php/gaussian

KETEPATAN KLASIFIKASI PEMILIHAN METODE KONTRASEPSI DI KOTA SEMARANG MENGGUNAKAN *BOOSTSTRAP AGGREGATTING* REGRESI LOGISTIK MULTINOMIAL

Ahmad Reza Aditya[1], Suparti[2], Sudarno[3]
[1]Mahasiswa Jurusan Statistika Undip
[2,3]Staf Pengajar Jurusan Statistika Undip

ABSTRACT

Classification is one of the statistical methods in grouping the data compiled systematically. Classification problem rises when there are a number of measures that consists of one or several categories that can not be identified directly but must use a measure. classification methods commonly used in studies to analyze a problem or event is logistic regression analysis. However, this classification method provides unstable parameter estimation. So to obtain a stable parameter multinomial logistic regression model used bootstrap approach that is bootstrap aggregating (bagging). The purpose of this study was to compare the accuracy of the classification multinomial logistic regression models and bootstrap aggragatting model using the data of family planning in Semarang. From the results of bagging multinomial logistic regression obtained classification accuracy in replication bootstrap most 50 times at 51%, this model is able to decrease the classification error of up to 2% compared to the multinomial logistic regression model with a classification accuracy of 49%.

Keywords: logistic regression, bootstrap aggregating, accuracy of classification

1. Pendahuluan

Pengklasifikasian merupakan salah satu metode statistik dalam pengelompokan suatu data yang disusun secara sistematis. Masalah klasifikasi muncul ketika terdapat sejumlah ukuran yang terdiri dari satu atau beberapa kategori yang tidak dapat diidentifikasikan secara langsung tetapi harus menggunakan suatu ukuran. Terdapat beberapa metode klasifikasi yang biasa digunakan dalam penelitian-penelitian untuk menganalisa suatu masalah atau kejadian. Salah satunya yaitu metode analisis regresi logistik. Pada regresi logistik akan diperoleh suatu model logistik yang digunakan untuk menjelaskan hubungan antara prediktor dan respon (yang bersifat dikotomus atau ada dua kategori/kelompok), serta untuk mengelompokkan obyek ke dalam salah satu dari dua kategori respon. Dalam perkembangannya, regresi logistik dapat juga digunakan untuk respon kategori lebih dari dua kelompok, yang dikenal dengan regresi logistik polikotomus/multinomial. Regresi logistik merupakan sebuah metode analisis statistik untuk menggambarkan hubungan antara variabel prediktor dengan variabel respon yang mempunyai dua atau lebih kategori dengan variabel prediktor yang menggunakan skala kategorik maupun interval (Hosmer dan Lemeshow, 1989). Regresi logistik membentuk persamaan atau fungsi dengan pendekatan *Maximum Likelihood*, yang memaksimalkan peluang pengklasifikasian objek yang diamati menjadi kategori yang sesuai kemudian mengubahnya menjadi koefisien regresi yang sederhana. Metode pengklasifikasian ini memberikan pendugaan parameter yang tidak stabil, artinya jika terdapat perubahan dalam data set menyebabkan perubahan yang signifikan pada model (Breiman, 1994). Untuk memperoleh parameter yang stabil pada model regresi logistik multinomial digunakan pendekatan bootstrap yaitu metode *Bootstrap Aggregating (Bagging)*. Bagging diperkenalkan oleh Breiman (1994) adalah metode untuk memperbaiki kekuatan prediksi dari beberapa penduga atau algoritma tertentu seperti regresi atau pohon klasifikasi. Metode yang dinyatakan terbaik dalam perbandingan antara regresi logistik multinomial dengan *Bagging* Regresi logistik multinomial biasanya adalah metode yang memiliki tingkat kesalahan klasifikasi (*missclassified*) lebih kecil. Tingkat kesalahan klasifikasi dapat diketahui dari hasil ketepatan prediksi masing-masing metode yang dibandingkan dengan data aktualnya. Untuk lebih jelasnya bagaimana Regresi Logistik Multinomial dan *Bootstrap Aggregating* Regresi Logistik Multinomial bekerja dan metode mana yang lebih baik dalam memprediksi maka metode ini coba diaplikasikan pada data keluarga di kota Semarang 2013. Data tersebut merupakan data sekunder yang diambil dari pemutakhiran data keluarga Januari 2013-Januari 2014 yang dilakukan oleh BKKBN Jawa Tengah.

2. Tinjauan Pustaka

2.1. Regresi Logistik

Regresi Logistik merupakan salah satu bagian dari Analisis Regresi, yang digunakan untuk memprediksi probabilitas kejadian suatu peristiwa, dengan mencocokkan data pada fungsi logit kurva logistik. Metode ini merupakan model linear umum yang digunakan untuk regresi binomial. Seperti analisis regresi pada umumnya, metode ini menggunakan beberapa variabel prediktor, baik numerik maupun kategori (Hosmer dan Lemeshow, 2000).

2.2. Regresi Logistik Multinomial

Dalam analisis data dimana variabel respon adalah nominal, digunakan suatu metode yang merupakan pengembangan dari regresi logistik dan dikenal sebagai regresi logistik nominal atau *nominal logistic regression*, sedangkan untuk variabel respon ordinal digunakan regresi logistik ordinal atau atau *ordinal logistic regression* (McCullagh & Nelder, 1983).

Regresi logistik multinomial (nominal dan ordinal) merupakan salah satu pendekatan pemodelan yang dapat digunakan untuk mendeskripsikan hubungan beberapa variabel prediktor X dengan suatu variabel respon multinomial (polytomous). Model regresi logistik nominal digunakan ketika tidak ada urutan di antara kategori respon. Satu kategori diantaranya dipilih sebagai kategori acuan. Persamaan regresi logistik multinomial (Hosmer dan Lemeshow, 2000) secara umur adalah sebagai berikut:

$$P(Y=j|x)=\pi_j(x)=\frac{\exp\left[g_j(x)\right]}{\sum_{k=0}^{r-1}\exp[g_k(x)]}$$

$$=\frac{\exp\,(\beta_{j0}+\beta_{j1}x_1+\beta_{j2}x_2+\cdots+\beta_{jp}x_p)}{\sum_{k=0}^{r-1}\exp\,(\beta_{k0}+\beta_{k1}x_1+\beta_{k2}x_2+\cdots+\beta_{kp}x_p)}$$

Keterangan:

$P(Y=j|x)$ = perluang bersyarat dari variabel respon Y untuk kategori ke *j* pada vektor *x, j* = 0,1,..,*r-1*

$\pi_j(x)$ = persamaan regresi logistik untuk variabel respon Y untuk kategori ke *j*

$g_j(x)$ = logit pada variabel responY untuk kategori ke *j*

x_m = nilai dari variabel penjelas ke-m, m=1,2,3,...,p

β_{jm} = koefisien/parameter model

Dimana vektor $\tilde{\beta}_0=0$ sehingga $g_0(x)=0$

$\tilde{\beta}_0$ = parameter model logit untuk variabel respon Y untuk kategori ke 0 $(\beta_{00},\beta_{01},\dots,\beta_{0p})$

Suatu variabel respon dengan r kategori akan membentuk r-1 persamaan logit, dimana masing-masing persamaan ini membentuk regresi logistik biner yang membandingkan suatu kelompok kategori terhadap referensi, yaitu sebagai berikut:

$$g_{r-1}(x) = ln\frac{P(Y=r-1|x)}{P(Y=0|x)} = ln\left(\frac{\pi_{r-1}(x)}{\pi_0(x)}\right) = \beta_{(r-1)0} + \beta_{(r-1)1}x_1 + \cdots + \beta_{(r-1)p}x_p$$

Secara umum, langkah-langkah yang dilakukan dalam analisis regresi logistik multinomial adalah :

1. Estimasi parameter regresi logistik multinomial.
2. Melakukan pengujian parameter secara simultan untuk mengetahui kecocokan model analisis tersebut.
3. Melakukan pengujian parameter secara parsial untuk mengetahui variabel prediktor yang paling berpengaruh dalam model tersebut.
4. Melakukan interpretasi terhadap nilai rasio kecenderungan yang terbentuk.

2.3. Estimasi Parameter

Dalam model regresi logistik, nilai harapan antar variabel respon tidak linier serta memiliki varian yang tidak sama sehingga penduga parameter $\tilde{\beta}$ diperoleh melalui metode Maximum Likelihood (Hosmer & Lemeshow, 2000). fungsi likelihood bersyarat untuk sampel sebanyak n observasi sebagai berikut:

$$l(B) = \prod_{i=1}^{n}[\,\pi_0(x_i)^{y_{0i}}\pi_1(x_i)^{y_{1i}}\pi_2(x_i)^{y_{2i}} \ldots \pi_{r-1}(x_i)^{y_{(r-1)i}}]$$

Dengan demikian maka fungsi log likelihoodnya adalah:

$$L(\beta) = \ln[l(B)]$$

$$L(\beta) = \ln\left[\prod_{i=1}^{n}[\,\pi_0(x_i)^{y_{0i}}\pi_1(x_i)^{y_{1i}}\pi_2(x_i)^{y_{2i}} \ldots \pi_{r-1}(x_i)^{y_{(r-1)i}}]\right]$$

Untuk mendapatkan nilai $\tilde{\beta}$ yang memaksimalkan $L(\beta)$ dilakukan dengan mengambil turunan pertama dan kedua terhadap $L(\beta)$. Nilai β dapat ditentukan, dengan metode iterasi Newton-Rapshon.

2.4. Pengujian Parameter

1. Pengujian parameter secara simultan

   Pengujian parameter secara simultan dilakukan untuk menguji peranan variabel prediktor dalam model secara bersama-sama (Hosmer & Lemeshow, 2000).

   H_0 : $\beta_{j1} = \beta_{j2} = \cdots = \beta_{jp} = 0$, artinya tidak ada pengaruh antara sekumpulan variabel prediktor dengan variabel respon

   H : minimal ada satu $\beta_{jm} \neq 0$, artinya minimal ada satu variabel prediktor yang berpengaruh terhadap variabel respon, m = 1,2,..,p

   Dengan statistik uji :

   $$G = -2ln\left[\frac{l_0}{l_k}\right]$$

   H_0 akan ditolak pada tingkat signifikansi α apabila nilai $G > X^2{}_{(v;\alpha)}$ atau p-value < α. Jika H_0 ditolak maka disimpulkan bahwa variabel prediktor secara bersama-sama atau secara keseluruhan mempengaruhi variabel respon.

2. Pengujian parameter secara individu
   Pengujian parameter secara individu dilakukan untuk menguji peranan variabel prediktor dalam model secara individu. Pengujian variabel dilakukan satu per satu menggunakan statistik uji Wald (Hosmer & Lemeshow, 2000).
   H_0 : $\beta_{jm} = 0$, artinya tidak ada pengaruh variabel prediktor ke-m terhadap variabel respon kategori ke-j
   H_1 : $\beta_{jm} \neq 0$, artinya ada pengaruh variabel prediktor ke-m terhadap variabel respon kategori ke-j

   $j = 0,1,2,\dots,r-1;\ m = 1,2,\dots,p$

   Statistik ujinya adalah
$$W = \left[\frac{\hat{\beta}_{jm}}{Se(\hat{\beta}_{jm})}\right]^2$$
   H_0 akan ditolak jika $W > X^2{}_{(1;\alpha)}$ atau p-value < α. Jika H_0 ditolak maka dapat disimpulkan bahwa β_{jm} signifikan.

2.5. Prosedur Klasifikasi

Untuk menilai kemampuan prosedur klasifikasi dalam memprediksi keanggotaan kelompok, biasa digunakan tingkat kesalahan klasifikasi yang biasa disebut *apperent error rate*. Bisa juga menggunakan sebaliknya, yaitu *apperent correct classification rate* atau tingkat ketepatan klasifikasi (Rencher, 2002). Misalkan variabel respon *Y* terdiri dari 3 kategori, maka
$$Apparent\ error\ rate = \frac{n_{12}+n_{13}+n_{21}+n_{23}+n_{31}+n_{32}}{n_1+n_2+n_3}$$

2.6. Bagging (*Bootstrap Aggregating*)

Bagging (*bootstrap aggregating*) adalah metode untuk menghasilkan beberapa versi prediktor dan menggunakannya untuk mendapatkan prediktor agregat. Rata-rata agregat atas beberapa prediktor digunakan ketika memprediksi hasil numerik. Prediktor-prediktor terbentuk dengan replikasi bootstrap pada data dan menggunakannya sebagai data baru.

Himpunan data (*data set*) £ terdiri dari $\{(x_n, y_n), n = 1, \dots, N\}$ dengan *y* dapat berupa klas label atau numerik respon. Jika input adalah *x* maka y diprediksi menggunakan $\varphi(\boldsymbol{x},£)$ dimana $\varphi(\boldsymbol{x},£)$ adalah parameter. Selanjutnya dilakukan pembootstrapan data asli sehingga replikasi bootstrap ke-k diperoleh parameter$\{\varphi(\boldsymbol{x},£_k)\}$. Replikasi bootstrap dilakukan sebanyak B kali sehingga $\{£^{(B)}\}$ dari £ dan dibentuk parameter $\{\varphi(\boldsymbol{x},£^{(B)})\}$. $\{£^{(B)}\}$ adalah resampling dengan pengembalian. Jika *y* merupakan data numerik, prosedur nyata untuk menggantikan $\varphi(\boldsymbol{x},£)$ dengan mengambil rata-rata dari $\{\varphi(\boldsymbol{x},£_k)\}$. Jika $\varphi(\boldsymbol{x},£)$ memprediksi sebuah kelas $j \in \{1, \dots, r\}$, maka salah satu metode dengan menggabungkan $\{\varphi(\boldsymbol{x},£_k)\}$ untuk memprediksi kelas. Ambil bootstrap sampel dengan pengulangan $\{£^{(B)}\}$ dari £ dan membentuk $\{\varphi(\boldsymbol{x},£^{(B)})\}$.
$$\varphi_B(x) = average_B\, \varphi(\boldsymbol{x}, £^{(B)})$$
Jika *y* merupakan kategorik, maka untuk menentukan kategorik dengan menggunakan $\{\varphi(\boldsymbol{x},£^{(B)})\}$ dari $\varphi_B(x)$.
Selanjutnya algoritma *bagging* untuk regresi logistik adalah sebagai berikut (Akbar, dkk, 2010) :

1. Mengambil sampel bootstrap sebanyak n dari *data set* $£$ dengan pengulangan sebanyak n. Pengambilan sampel sedemikian hingga setiap variabel aggregate dalam setiap observasi.
2. Memodelkan regresi logistik hasil sampel bootstrap $£^{(B)}$.
3. Menghitung peluang kumulatif, peluang masing-masing kategori respon untuk setiap observasi dan menghitung ketepatan klasifikasi. Kesalahan klasifikasi pada langkah ini disebut e_B.
4. Mengulang langkah 1 - 3 sebanyak B kali (Replikasi bootstrap).
5. Memperoleh ketepatan klasifikasi *bagging* yaitu rata-rata ketepatan klasifikasi setiap pengulangan sampai B. Sehingga kesalahan klasifikasi *bagging* untuk replikasi B kali adalah $\bar{e}_B$.
6. Membentuk model *bagging* regresi logistik dari rata-rata setiap parameter pada pengulangan sampai B.

3. Metodelogi Penelitian

3.1. Sumber Data

Data yang digunakan pada penelitian ini adalah data Keluarga Berencana di kota Semarang 2013. Data tersebut merupakan data sekunder yang diambil dari pemutakhiran data keluarga Januari 2013-Januari 2014 yang dilakukan oleh BKKBN Jawa Tengah. Jumlah data yang digunakan sebanyak 288 data.

3.2. Variabel Penelitian

1. Variabel Respon

Dalam permasalahan ini variabel respon yang digunakan adalah metode KB yang tersedia. Dimana Metode KB dibagi menjadi 7 metode yaitu IUD, Kondom, Suntikan, MOW, MOP, Implant, dan Pil. Pengkategorian untuk variabel respon adalah 1 = IUD, 2 = Kondom, 3 = Suntikan, 4 = MOW, 5 = MOP, 6 = Implant, dan 7 = Pil.

2. Variabel Prediktor

Tabel 1. Variabel Prediktor

Variabel Prediktor	Definisi	Kategori (Koding)	Jenis Variabel
X_1	Usia Suami	≤ 30 tahun (0) >30 tahun (1)	Kategorik
X_2	Usia Istri	≤ 30 tahun (0) >30 tahun (1)	Kategorik
X_3	Umur Anak Terakhir	≤ 5 tahun (0) > 5 tahun (1)	Kategorik
X_4	Jumlah Anak	≤ 2 anak (0) > 2 anak (1)	Kategorik
X_5	Pekerjaan Istri	Tidak terampil (0): Petani, Nelayan, Pedagang, Lainnya/Pekerja Lepas,	Kategorik

X_6	Pekerjaan Suami	Tidak Bekerja Terampil (1): Pegawai Negeri, TNI/POLRI, Pegawai Swasta, Wiraswasta, Pensiunan	Kategotik
X_7	Pendidikan Suami	SD (1) SMP (2) SMA (3) Perguruan Tinggi (4)	Kategorik
X_8	Pendidikan Istri		Kategotik
X_9	Tingkat Kesejahteraan Keluarga	KS 1 (1) KS 2 (2) KS 3 dan KS 3+ (3)	Kategorik

3.3. Metode Analisis

Metode analisis yang akan digunakan adalah *Bagging* Regresi Logistik Multinomial. Adapun tahap-tahap dalam *Bagging* Regresi Logistik Multinomial adalah sebagai berikut.

1. Melakukan analisis regresi logistik multinomial dengan mengestimasi parameter dan melakukan pengujian secara bersama dan secara individu terhadap masing-masing variabel prediktor.
2. Mendapatkan variabel prediktor yang signifikan berpengaruh terhadap model regresi logistik multinomial.
3. Menentukan model regresi logistik multinomial dengan memasukkan seluruh variabel prediktor yang signifikan berpengaruh pada pengujian secara individu.
4. Menentukan kesalahan klasifikasi regresi logistik multinomial.
5. Melakukan bootstrap aggregating untuk prediktor dari model logistik multinomial, sebanyak B(50, 60, 70, 80, 90, 100, 150, dan 200) replikasi bootstrap.
6. Menentukan ketepatan klasifikasi pada setiap pengambilan sampel B replikasi bootstrap, sehingga diperoleh kesalahan klasifikasi e_B.
7. Menentukan rata-rata kesalahan klasifikasi bagging B.
8. Membandingkan klasifikasi model regresi logistik multinomial dan bagging regresi logistik multinomial.

4. Analisis dan Pembahasan

4.1. Model Awal Regresi Logistik Multinomial

Adapun persamaan logit yang diperoleh adalah sebagai berikut.

Logit 1

$$g_1(x) = 1{,}694 - 1{,}251x_1 - 0{,}824x_2 + 0{,}959x_3 - 0{,}233x_4 + 0{,}575x_5 - 0{,}694x_6 - 1{,}992x_{7(1)} - 0{,}909x_{7(2)} - 0{,}222x_{7(3)} - 1{,}707x_{8(1)} - 1{,}406x_{8(2)} - 1{,}696x_{8(3)} + 0{,}055x_{9(1)} + 0{,}455x_{9(2)}$$

Logit 2

$$g_2(x) = 2{,}175 + 1{,}371x_1 - 1{,}203x_2 + 0{,}299x_3 - 0{,}332x_4 + 0{,}392x_5 - 0{,}856x_6 - 3{,}486x_{7(1)} - 2{,}123x_{7(2)} - 1{,}192x_{7(3)} + 0{,}851x_{8(1)} - 0{,}632x_{8(2)} - 1{,}481x_{8(3)} - 0{,}584x_{9(1)} - 0{,}843x_{9(2)}$$

Logit 3

$$g_3(x) = 1{,}029 + 0{,}466x_1 - 0{,}602x_2 + 0{,}437x_3 - 0{,}072x_4 + 0{,}376x_5 - 0{,}254x_6 - 1{,}852x_{7(1)} + 0{,}187x_{7(2)} + 0{,}275x_{7(3)} + 0{,}471x_{8(1)} - 1{,}071x_{8(2)} - 1{,}478x_{8(3)} + 1{,}361x_{9(1)} + 1{,}830x_{9(2)}$$

Logit 4

$$g_4(x) = 2{,}767 - 0{,}342x_1 - 0{,}755x_2 - 0{,}133x_3 - 1{,}837x_4 - 0{,}297x_5 + 0{,}260x_6 + 0{,}075x_{7(1)} + 0{,}008x_{7(2)} - 0{,}506x_{7(3)} - 1{,}201x_{8(1)} - 1{,}357x_{8(2)} - 1{,}187x_{8(3)} - 0{,}213x_{9(1)} - 0{,}622x_{9(2)}$$

Logit 5

$$g_1(x) = -34{,}665 + 0{,}271x_1 - 15{,}980x_2 - 17{,}895x_3 + 17{,}361x_4 - 18{,}697x_5 + 14{,}332x_6 - 17{,}749x_{7(1)} - 21{,}505x_{7(2)} - 2{,}604x_{7(3)} + 21{,}362x_{8(1)} + 21{,}451x_{8(2)} + 18{,}960x_{8(3)} - 19{,}425x_{9(1)} - 16{,}831x_{9(2)}$$

Logit 6

$$g_6(x) = 1{,}735 - 18{,}805x_1 - 1{,}059x_2 + 0{,}484x_3 - 0{,}795x_4 + 0{,}666x_5 - 0{,}220x_6 - 4{,}612x_{7(1)} - 2{,}495x_{7(2)} - 1{,}340x_{7(3)} + 1{,}978x_{8(1)} - 1{,}020x_{8(2)} - 1{,}381x_{8(3)} + 0{,}340x_{9(1)} - x_{9(2)}$$

1. Uji Parameter Secara Bersama
   Hipotesis yang digunakan:
   H_0 : $\beta_{11} = \beta_{12} = \cdots = \beta_{jp} = 0$, artinya tidak ada pengaruh antara variabel prediktor dengan variabel respon (model tidak signifikan).
   H_1 : minimal ada satu $\beta_{jm} \neq 0$, artinya minimal ada satu variabel prediktor yang berpengaruh terhadap variabel respon (model signifikan).
   Statistik uji likelihood G sebesar 174,221 dengan signifikansi sebesar 0,000. Kemudian nilai tersebut dibandingkan dengan tabel distribusi chi kuadrat dengan derajat bebas 84. Nilai chi kuadrat dengan derajat bebas 84 dan tingkat signifikansi α = 5% adalah 106,40. Kriteria uji untuk uji likelihood, H_0akan ditolak pada tingkat signifikansi α apabila nilai $G > X^2{}_{(v;\alpha)}$ atau p-value < α. Berdasarkan kriteria tersebut maka diperoleh kesimpulan bahwa model awal signifikan karena nilai $G = 140{,}282 > X^2{}_{(84;0{,}05)} = 106{,}40$ dan p-value = 0,000 < α = 5%.

2. Pengujian Parameter Secara Individu Model Awal
   Hipotesis yang digunakan adalah
   H_0 : $\beta_{j1} = 0$ (Koefisien tidak signifikan).
   H_1 : $\beta_{j1} \neq 0$ (Koefisien signifikan).
   Taraf signifikansi yang digunakan adalah α = 5%.
   Kriteria uji yang digunakan adalah H_0 akan ditolak jika $W > X^2{}_{(1;\alpha)}$ atau p-value < α

Tabel 2. Hasil Uji Wald Model Awal

Variabel Prediktor	Definisi	Hasil Uji Wald	Kesimpulan
X_1	Usia Suami	Tidak ada Koefisien yang signifikan	variabel X_1 dikeluarkan dari model
X_2	Usia Istri	Tidak ada Koefisien yang signifikan	variabel X_2 dikeluarkan dari model
X_3	Umur Anak Terakhir	Tidak ada Koefisien yang signifikan	variabel X_3 dikeluarkan dari model
X_4	Jumlah Anak	Ada koefisien yang	variabel X_4 tetap

		signifikan	di model
X_5	Pekerjaan Istri	Tidak ada Koefisien yang signifikan	variabel X_5 dikeluarkan dari model
X_6	Pekerjaan Suami	Tidak ada Koefisien yang signifikan	variabel X_6 dikeluarkan dari model
X_7	Pendidikan Suami	Ada koefisien yang signifikan	variabel X_7 tetap di model
X_8	Pendidikan Istri	Ada koefisien yang signifikan	variabel X_8 tetap di model
X_9	Tingkat Kesejahteraan Keluarga	Ada koefisien yang signifikan	variabel X_9 tetap di model

4.2. Model Kedua Regresi Logistik Multinomial

Adapun persamaan logit yang diperoleh adalah sebagai berikut.

Logit 1

$$g_1(x) = 2{,}433 - 0{,}559x_4 - 2{,}167x_{7(1)} - 1{,}131x_{7(2)} - 0{,}424x_{7(3)} - 1{,}589x_{8(1)} - 1{,}330x_{8(2)} - 1{,}627x_{8(3)} + 0{,}213x_{9(1)} + 0{,}611x_{9(2)}$$

Logit 2

$$g_2(x) = 2{,}476 - 0{,}343x_4 - 3{,}711x_{7(1)} - 2{,}524x_{7(2)} - 1{,}269x_{7(3)} + 0{,}978x_{8(1)} - 0{,}434x_{8(2)} - 1{,}392x_{8(3)} - 0{,}608x_{9(1)} - 0{,}912x_{9(2)}$$

Logit 3

$$g_3(x) = 1{,}406 - 0{,}204x_4 - 1{,}934x_{7(1)} + 0{,}063x_{7(2)} + 0{,}201x_{7(3)} + 0{,}552x_{8(1)} - 0{,}980x_{8(2)} - 1{,}397x_{8(3)} + 1{,}429x_{9(1)} + 1{,}873x_{9(2)}$$

Logit 4

$$g_4(x) = 2{,}668 - 1{,}975x_4 - 0{,}031x_{7(1)} + 0{,}105x_{7(2)} - 0{,}578x_{7(3)} - 1{,}067x_{8(1)} - 1{,}414x_{8(2)} - 1{,}238x_{8(3)} - 0{,}298x_{9(1)} - 0{,}762x_{9(2)}$$

Logit 5

$$g_1(x) = -35{,}397 + 17{,}314x_4 - 19{,}902x_{7(1)} - 19{,}749x_{7(2)} - 1{,}916x_{7(3)} + 19{,}608x_{8(1)} + 17{,}797x_{8(2)} + 17{,}905x_{8(3)} - 17{,}849x_{9(1)} - 17{,}437x_{9(2)}$$

Logit 6

$$g_6(x) = 2{,}192 - 1{,}108x_4 - 4{,}885x_{7(1)} - 2{,}757x_{7(2)} - 1{,}522x_{7(3)} + 2{,}400x_{8(1)} - 0{,}764x_{8(2)} - 1{,}170x_{8(3)} + 0{,}413x_{9(1)} - 0{,}911x_{9(2)}$$

1. Uji Parameter Secara Bersama

   Hipotesis yang digunakan:

   H_0 : $\beta_{11} = \beta_{12} = \cdots = \beta_{jp} = 0$, artinya tidak ada pengaruh antara variabel prediktor dengan variabel respon (model tidak signifikan).

   H_1 : minimal ada satu $\beta_{jm} \neq 0$, artinya minimal ada satu variabel prediktor yang berpengaruh terhadap variabel respon (model signifikan).

   Statistik uji likelihood sebesar 140,282 dengan signifikansi sebesar 0,000. Kemudian nilai tersebut dibandingkan dengan tabel distribusi chi kuadrat dengan derajat bebas 54. Nilai chi kuadrat dengan derajat bebas dan tingat signifikansi α =5% adalah 72,15. Kriteria uji untuk uji likelihood, H_0 akan ditolak pada tingkat signifikansi α apabila nilai $G > X^2{}_{(v;\alpha)}$ atau p-value < α. Berdasarkan kriteria tersebut maka diperoleh kesimpulan bahwa model awal signifikan karena nilai $G = 140{,}282 > X^2{}_{(54;0{,}05)} = 72{,}15$ dan p-value = 0,000 < α = 5%.

2. Pengujian Parameter Secara Individu Model Kedua
   Hipotesis yang digunakan adalah
   $H_0 : \beta_{j1} = 0$ (Koefisien tidak signifikan).
   $H_1 : \beta_{j1} \neq 0$ (Koefisien signifikan).
   Taraf signifikansi yang digunakan adalah α = 5%.
   Kriteria uji yang digunakan adalah H_0 akan ditolak jika $W > X^2{}_{(1;\alpha)}$ atau p-value < α

Tabel 3. Hasil Uji Wald Model Kedua

Variabel Prediktor	Definisi	Hasil Uji Wald	Kesimpulan
X_4	Jumlah Anak	Ada koefisien yang signifikan	variabel X_4 tetap di model
X_7	Pendidikan Suami	Ada koefisien yang signifikan	variabel X_7 tetap di model
X_8	Pendidikan Istri	Ada koefisien yang signifikan	variabel X_8 tetap di model
X_9	Tingkat Kesejahteraan Keluarga	Ada koefisien yang signifikan	variabel X_9 tetap di model

Karena semua variabel prediktor sudah signifikan maka model ini dapat digunakan dan digunakan sebagai model akhir.

Persentase tingkat ketepatan klasifikasi untuk model regresi logistik multinomial adalah sebesar 49%, sehingga untuk persentase tingkat kesalahan klasifikasi sebesar $1 - apper =$ 51%.

4.3. *Bagging* Regresi Logistik Multinomial

Sampel bootstrap diambil sebanyak n data yaitu 288 data, kemudian direplikasi bootstrap sebanyak 50, 60, 70, 80, 90, 100, 150 dan 200. Pada setiap pengambilan sampel akan dibentuk model regresi logistik multinomial sehingga akan diperoleh nilai ketepatan klasifikasi sebanyak B dalam setiap B replikasi bootstrap.

Tabel 4. Ketepatan Klasifikasi *Bagging* Regresi Logistik

Replikasi bootstrap	Ketepatan Klasifikasi	$\bar{e}_B$	e	Penurunan kesalahan klasifikasi
50	51,0%	49,0%	51,0%	2%
60	50,3%	49,7%	51,0%	1,3%
70	50,8%	49,2%	51,0%	1,8%
80	50,4%	49,6%	51,0%	1,4%
90	50,5%	49,5%	51,0%	1,5%
100	49,9%	50,1%	51,0%	0,9%
150	50,4%	49,6%	51,0%	1,4%
200	50,4%	49,6%	51,0%	1,4%

Tabel 4 memberikan informasi bahwa dengan 50 replikasi bootstrap diperoleh rata- rata ketepatan klasifikasi terbesar yaitu sebesar 51%, sehingga berdasarkan hasil pada Tabel 4, maka dapat disimpulkan bahwa diperoleh bagging prediktor terbaik adalah pada replikasi bootstrap sebanyak 50 kali. Model bagging ini dapat meningkatkan ketepatan klasifikasi dari model data set tunggal yaitu sebesar 49% menjadi 51% atau dengan kata lain bagging dapat menurunkan kesalahan klasifikasi sebesar 2% dari model data set tunggal.

5. Kesimpulan

1. Model akhir dibentuk dengan parameter yang berpengaruh terhadap pemilihan metode kontrasepsi (Y) yaitu Jumlah Anak (X_4), Pendidikan Suami (X_7), Pendidikan Istri (X_8), dan Tingkat Kesejahteraan Keluarga (X_9).
2. Perbandingan ketepatan klasifikasi model regresi logistik multinomial dengan model bagging regresi logistik multinomial diperoleh bahwa model terbaik adalah model bagging regresi logistik multinomial dengan replikasi sebanyak 50 kali dengan tingkat ketepatan klasifikasi terbesar, yaitu 51%.
3. Penurunan persentase tingkat kesalahan model bagging dengan 50 kali replikasi sebesar 2% merupakan penurunan persentase tingkat kesalahan terbesar.

Daftar Pustaka

Agresti, A. 1990. *Categorical Data Analysis*. John Wiley and Sons. New York.

Akbar, M.S., Adatul M.& Lalita P. 2010. *Klasifikasi Status Gizi Balita Dengan Bagging Regresi Logistik Ordinal* (Studi Kasus Survey Kekurangan Energi Protein Kabupaten Nganjuk).Media Statistika. Volume 3.

Breiman, L. 1994. *Bagging Predictor*. Technical report No. 421. Departement of statistics University of California.

Efron, B., dan R. J. Tibshirani. 1993. *An Introduction to the Bootstrap*. Chapman and Hall. New York.

Hosmer, D.W., dan Lemenshow. 2000. *Applied Logistic Regression*. John Wiley and Sons.USA.

McCullagh, P. dan J.A. Nelder. 1989. *Generalized Linear Models*. 2nd ed. Chapman & Hall/CRC, Boca Raton, Florida.

Rencher, A.C. 2002. *Methods of Multivariate Analysis*. Second edition. John Wiley & Son. USA.

Sulistio, E. dan Ispriyanti, D. 2010.*Penerapan Regresi Logistik Multinomial Pada Pemilihan Alat Kontrasepsi Wanita* (Studi Kasus Di Desa Tonggara Kecamatan Kedungbanteng Kabupaten Tegal). Media Statistika. Volume 3.